Oulu University - B.Sc. Thesis
The period luminosity relation of Cepheids in the light of Gaia

Markus Mikael Anetjärvi -2459934

Summer 2020

Abstract

Cepheids are bright variable stars that pulsate radially and have a strong relation between the luminosity and the period of the pulsation. The relation is known as the Period-Luminosity (PL) relation making them useful for extragalactic distance estimations.

The goal of this Thesis is to determine the PL/PW relations of Cepheids using the Gaia data, that contains observations of 9575 individual Cepheids. The PL/PW relations are calibrated using the Magellanic Clouds, and the results are consistent with the literature values. The calibrated relations are then used to the Milky Way Cepheids, and distances obtained by the relations are well consistent with the Gaia parallax distances, up to 3000 parsecs.

Contents

1 Introduction

1.1 Cepheids as the rung of the Cosmic Distance Ladder

Measuring distances to celestial objects has always been one of the fundamental and challenging goals in astronomy. One example of brilliant distance estimation methods is from ancient Greece astronomer, Aristarchus of Samos, as he was able to estimate the relative distances to the Sun and Moon and their sizes using geometry.[1][2] To make estimations of distances and sizes, he used Earth's size as well as solar and lunar eclipses.[2] For example, he measured during a lunar eclipse that the diameter of Earth's shadow is approximately double the size of the Moon.[1] Results obtained by Aristarchus of Samos differ from the values that are known nowadays, which is understandable due to the increased accuracy of measurement instruments. Even though the results are wrong, the methods are correct.

Today, astronomers have high accuracy instruments that have allowed them to estimate distances to the distant universe. However, estimating distances to distant objects in astronomy is not a simple task, but requires different methods for different distances. These methods together form the so-called Cosmic distance ladders. The main parts, rungs, of the Cosmic distance ladders are explained in this section while focusing more on the Cepheids.

One of the first rungs is a method called trigonometric parallax.[3] The parallax angle is the angular displacement of the object compared to the background stars, while the observer moves. So the parallax method is a geometric distance estimation method, and in Figure 1, one sees how the parallax method is applied in astronomy and in the Gaia mission.

After the parallax angle has been measured, the distance to the object can be calculated by basic trigonometry, as seen in Equation 1.

$$r = \frac{1}{\varpi} \quad \text{,when r} >> 1 \text{ AU} \tag{1}$$

Where ω is the measured parallax in arcseconds, and r is the distance to the object in parsecs[1]. Before the Gaia mission[5], parallaxes could be measured accurately only for nearby stars, up to hundreds of parsecs. Now with Gaia[5], accurate parallaxes can be measured up to thousands of parsecs, as will be seen in Section 4.1 Results.

When the observed star is too far away, one can not measure the parallax angle accurately anymore, and so one needs to move to the next rung of the Cosmic distance ladders. Using measured parallax distances and apparent magnitudes of nearby stars, one can compute the stars' absolute magnitude[2] —see Equation 6—, and if one plots stars' absolute magnitude against the measured color of stars, one gets the Hertzprung-Russel (HR) diagram[3] seen in Figure 2.

[1]Parsec is the distance at which 1 AU can be seen in 1-arcsecond angle. 1 pc = 3.26 ly = 206264.806 AU

[2]Absolute magnitude M is the star's magnitude if it would be at the distance of 10 pc.[6]

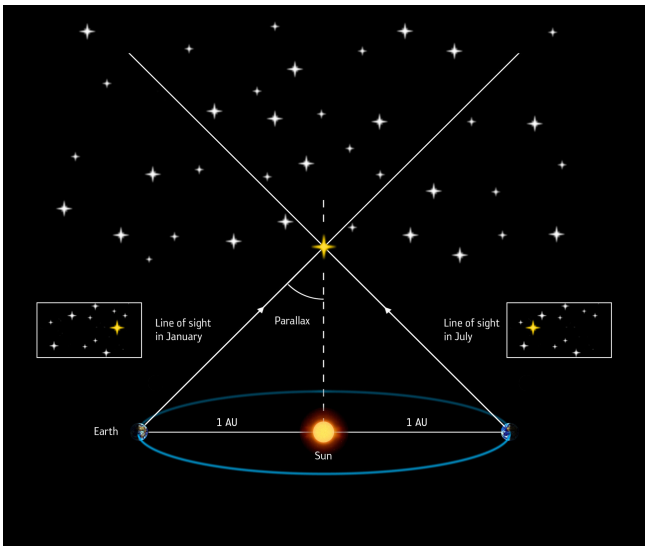

Figure 1: Demonstration figure for measuring annual parallax. The angular displacement of the object compared to the reference background during the half-year period is called the annual parallax. Picture is from "ESA Gaia mission, ESA/ATG medialab[4]"

From the HR diagram, one can obtain the star's absolute magnitude when the color and type of star are known. Now when one has the measured apparent magnitude and estimated absolute magnitude from the HR diagram, one can get the distance using distance modulus, see Equation 6. This method is called the main sequence fitting, and it is often used to estimate the distance to the clusters in the Milky Way, as they provide many stars approximately at the same distance. As the name suggests, the main sequence fitting works best for the stars in the main sequence, as their color-brightness relation is tightest, as seen in Figure 2. However, the HR diagram provides more distance estimation methods, such as the red giant branch method.[3] The main sequence fitting works far further than the parallax method, up to tens of thousands parsecs.[8]

As the distance gets larger, one can not measure the color and magnitude of individual main sequence stars anymore because they are not bright enough, and the resolution of instruments is limited. The next rung in the cosmic distance ladder is variable stars and especially Cepheid variables. Cepheid variables are bright giant stars that oscillate in brightness periodically[3], which can be seen in Figure 3. Figure 3 represents the light curve of Delta Cepheid, and indeed it can be seen that the brightness oscillates periodically.

In the early 20th century, Harvard astronomer Henrietta Swan Leavitt (1868-1921)[8] made the important observation of Cepheid variables in the Magellanic Clouds, and by assuming that the Cepheids are at the same distances[3] in Magel-

[3]Distances to Magellanic Clouds are much larger than the size of either galaxy, and so

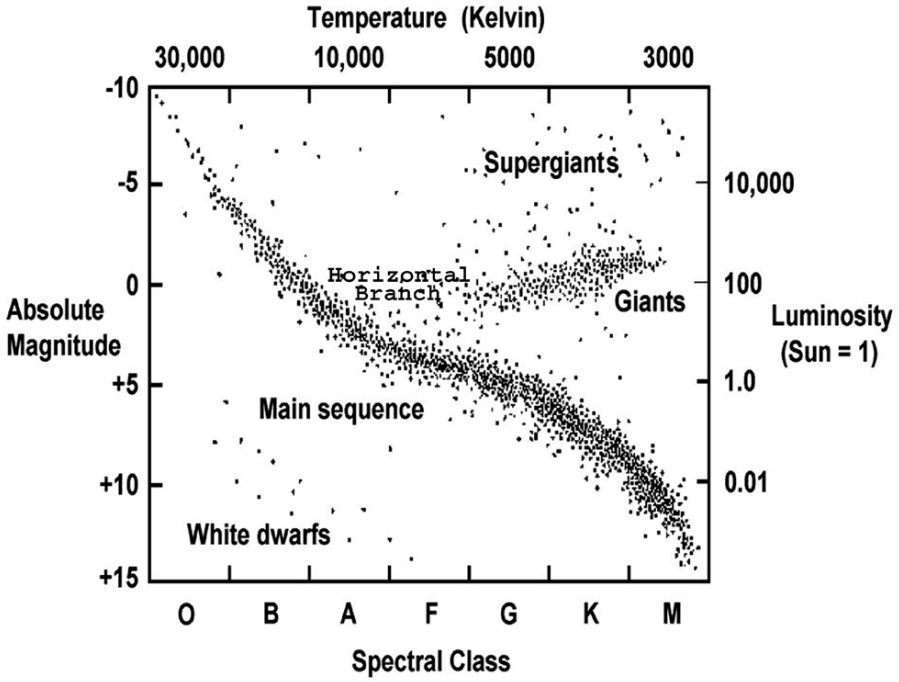

Figure 2: The Hertzprung-Russel (HR) Diagram. Different phases of stars' evolution can be seen in the diagram. Diagram is from "Chandra X-Ray Observatory, NASA/CXC/SAO[7]"

lanic Clouds, she could plot Cepheids' brightness against the period. She noticed that the Cepheids with a longer pulsating period are brighter than those with a shorter period. She could conclude that they vary in intrinsic brightness. However, the measured brightness' of Cepheids in the Magellanic Clouds were in apparent magnitudes, and thus they can not be used to distance estimations yet. Cepheids were also observed in nearby clusters, and the distance to the clusters was estimated using the main sequence fitting described before.[8] Using the known distance to Cepheids, astronomers were able to calibrate the Period-Luminosity relation, and thus the distance to Magellanic Clouds could be solved. The distances to the Magellanic Clouds are nowadays confirmed by multiple different methods[4][10], and the obtained distances are used to calibrate the Period-Luminosity relation of Cepheids as seen in the following Sections.

Cepheids are common stars, and so they can be spotted in nearby galaxies. Distances to those nearby galaxies, up to hundreds of thousands of parsecs, can be then estimated using the calibrated Period-Luminosity relation of Cepheids.

every Cepheid within Large Magellanic Cloud or Small Magellanic Cloud can be thought to be at the same distance.

[4] i.e. using eclipsing binary stars[10]

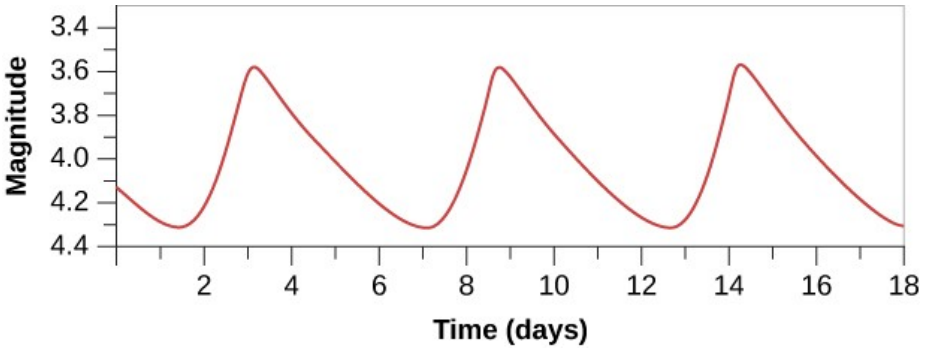

Figure 3: The light curve of the Delta Cepheid and periodic oscillation of brightness can be seen. Picture is from "Lumen learning[9]"

During the year 1923, Edwin Hubble (1889-1953)[8] found a Cepheid variable in the Andromeda galaxy.[11] Using the known Period-Luminosity relation of Cepheids, Hubble proved that the Andromeda galaxy is not part of the Milky Way galaxy, as was commonly thought in those days. By that observation of a single variable, the size of the universe known expanded dramatically.[3]

Cepheids are relatively young stars, and so they can not be found in older stellar population zones, but are instead found in active star-forming zones, such as the disk plane of the galaxy.[3] In Figure 4, one sees the Cepheids in the galactic coordinate system measured by Gaia. One indeed notices that most of the Cepheids in the Milky Way are located in the disk plane. The Magellanic Clouds can also be seen in Figure 4. The disk plane in the Milky Way contains a lot of dust, which causes extinction, reducing the measured apparent brightness. The effect of dust extinction need to be taken into account, and it is discussed in Section 2.3.

Even though the Cepheids are bright, they can not be seen further than hundreds of thousands of parces, and the next rung of the Cosmic distance ladder is required. One of the brightest events in the universe, supernovae, can be seen from extreme distances. There exist multiple supernovae types, but the type 1a supernova is the most important in terms of standard candles[5].[3] A type 1a supernova occurs in a binary systems which includes a white dwarf and a donor companion star. Material flows from the companion star to the white dwarf, increasing its mass. After the white dwarf reaches the Chandrasekhar limit, 1.4 solar masses, the white dwarf becomes unstable and explodes as a supernova.[12] Due to this mechanism, the type 1a supernovae release roughly the same amount of energy when they explode, making them excellent standard candles. Explosion of the supernova emits so much light, that it can outshine its whole host galaxy making it visible from distant galaxies, up to hundreds of

[5]Standard candles in astronomy are sources with well known absolute magnitude.[3] Distance to the standard candle can be solved using distance modulus —see Equation 6—, when the absolute magnitude is known, and the apparent magnitude is measured.

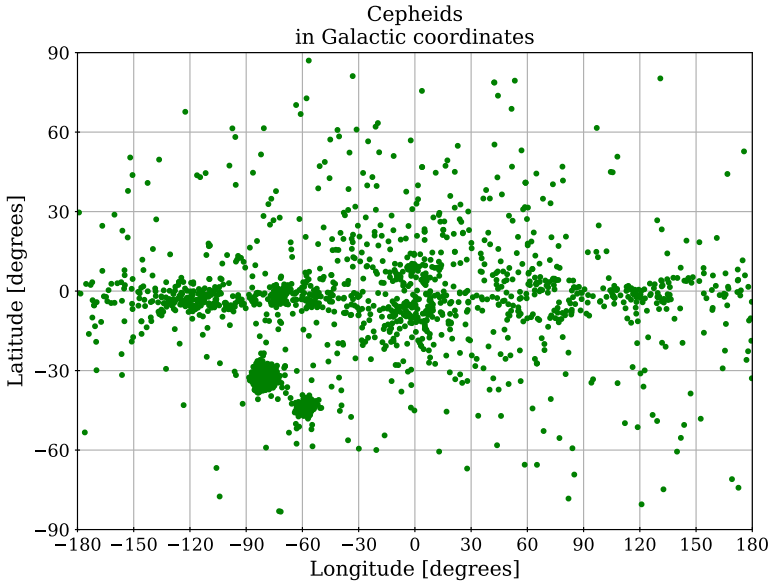

Figure 4: Cepheids observed by the Gaia in Galactic coordinates. One sees the disk plane of the Milky Way, as well as the Large and Small Magellanic Clouds. Most of the Cepheids are located in the disk plane, but not every. Cepheids that are not in the disk plane might be misclassified objects. Misclassification is discussed in Section 2.1.

millions of parsecs.[12] When a type 1a supernova has happened in a neighboring galaxy, one can use the Cepheids as distance estimation to calibrate the absolute brightness of the type 1a supernova.

Distances to the most distant galaxies are obtained by measuring the redshift of the galaxy and this distance estimation method related to redshift is the last rung of the Cosmic distance ladder. The measured spectrum of the galaxy's light becomes redshifted as the galaxy moves further away from the Milky Way.[3] Hubble's law states that the further the galaxy is, the faster it is receding as the universe is expanding.[13] Hubble's law[3] is written as:

$$v_r = H_0 d \tag{2}$$

Where v_r is the receding velocity, H_0 is the Hubble constant, and d is the distance to the galaxy. By measuring the galaxy's redshift, one gets the receding speed of the galaxy, and by knowing the velocity, one can use Equation 2 to calculate how far the galaxy is. Distance results from previous rungs have been used to calibrate the value of Hubble constant H_0.

Figure 5 shows the basic rungs of the Cosmic distance ladder and how far one can use them. As a summary of the Cosmic distance ladder, one needs

to use different distance estimation methods depending on how far the object is. As in regular ladders, the previous rung supports the next rung, and so the distance results from previous rungs are used to calibrate next rungs in the Cosmic distance ladder. Other rungs, i.e., RR-Lyrae variables and Tully-Fisher relation exist, and they are used similarly to estimate distances and calibration purposes.[3] The methods in the Cosmic distance ladder, especially a type 1a supernova, have also been used to discover that the expansion of the universe is accelerating.[3]

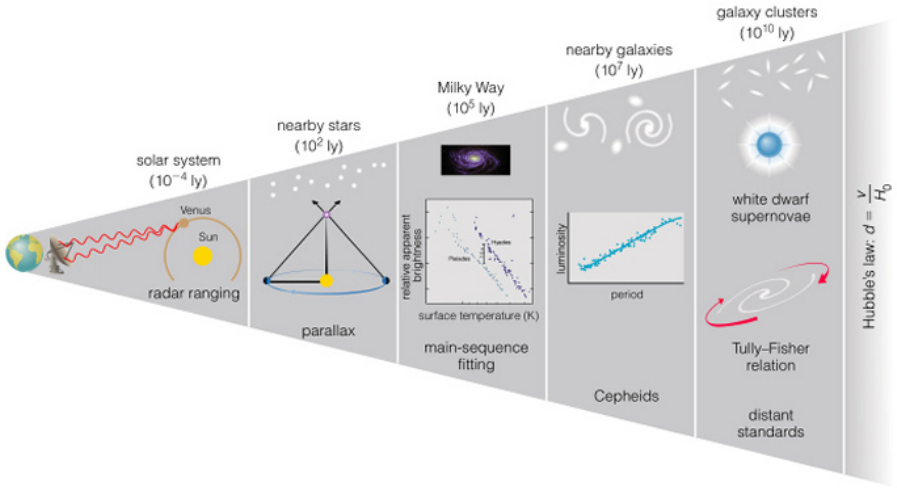

Figure 5: The Cosmic distance ladder. Some of the basic distance estimation methods are shown, as well as how far one can use that specific method. Picture is taken from "Cosmic distance ladder" by David Darling[14].

1.2 Before and after the Gaia mission

In the year 1989, astronomy took a massive leap to the secrets of the Milky Way, as the Hipparcos mission was launched by European Space Agency (ESA).[15] During its 3.5 years of observing, Hipparcos took accurate measurements for more than 100 000 sources and less accurate measurements for more than 1 million objects. Such a great amount of objects made it the largest astronomical data set on that time.[16] The Hipparcos mission was also the first space mission, whose primary focus was on measuring the position and distances to the stars. The Hipparcos mission allowed astronomers to study such a large number of stars in the Milky Way for the first time and provided research for years.[16]

In the year 2013, the Gaia mission by ESA was launched.[5] Gaia follows the steps of Hipparcos, as it also measures the positions and distances of the stars, but doing it 200 times more accurately.[17] Gaia does not only measure the astrometric properties of stars but also measures photometric and spectroscopic

properties. In Figure 6 is shown what Gaia has measured, and released in the second data release. The number of million measured positions of stars from the Hipparcos satellite now increased to over 1.5 billion stars measured by the Gaia. The Gaia has already measured an enormous amount of data, over 80 Terabytes, and is still measuring more.[18]

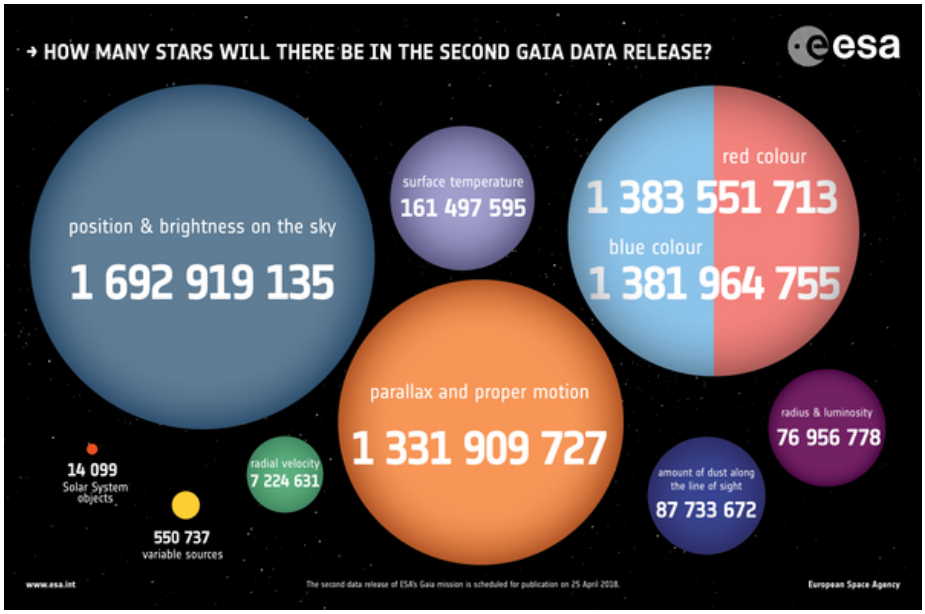

Figure 6: Measured properties of the second data release by the Gaia.Picture is from Gaia's webpage [19].

The data measured by Gaia will be used in various research projects. The list of science topics and goals can be found from Gaia's science webpage [20]. The scientific goal of this Thesis is to determine magnitude-period relation for the Cepheids measured by the Gaia.

2 Data processing

As previously mentioned, the Gaia satellite has already measured an enormous amount of data during its mission, over 80 Terabytes of purely scientific data.[18] In order to identify variable stars, Gaia observers each source about 70 times[21] so that time series analysis can be done. Using multiple algorithms and pipelines to analyze the data, Gaia can classify variable stars into specific variable classes.[22] Full list of the processing steps, variable classes, and classification models are found in Gaia Data Release 2- documentation [23][22].

A large volume of data brings its challenges, and rarely anyone needs all the specific data gathered by Gaia. Thus it is recommended to select the data one

needs carefully. The data can be found, selected, and downloaded from Gaia archive [24] as has been done in this Thesis. The Gaia archive uses Astronomical Data Query Language (ADQL) as a programming language, and multiple query examples can be found in the Gaia archive. See Appendix A for queries used in this Thesis.

While selecting the necessary data for this Thesis, inconsistencies arose between specific data columns. Inconsistencies are briefly discussed, and parallax and Wesenheit magnitude are described and explained in the next sections, before starting the analysis.

2.1 Data inconsistencies

The Gaia Data Release 2 (DR2)[24] has multiple data tables for different purposes. The data tables used in this Thesis are *gaiadr2.vari_cepheid*, which describe fundamental properties of Cepheid stars, such as period, magnitudes, and classification; *gaiadr2.vari_classifier_result*, which contains only information of classification and its confidence level for variables; and *gaiadr2.gaia_source*, which has information such as parallax, coordinates, and magnitudes for every source observed by Gaia.[24] The data of this Thesis have been divided into two cases, Case1 and Case2, for the reason that is described below.

Each source observed by Gaia has its unique source identifier, called *source_id*. The two data tables *gaiadr2.vari_cepheid* and *gaiadr2.vari_classifier_rersult* both have their classification column for the object sharing the *source_id*. The data of Case1 is based on the object classification of *gaiadr2.vari_cepheid* and is combined with *gaiadr2.vari_cepheid* and *gaiadr2.gaia_source* data tables based on the matching *source_id*. The data of Case2 is based on the object classification of *gaiadr2.vari_classifier_result* and is combined with *gaiadr2.vari_cepheid* and *gaiadr2.gaia_source* in order to get all necessary data for Cepheid variables.

	Case1		**Case2**
Column	'type_best_classification'		'Best_class_name'
	DCEP: 8890		CEP: 5478
	T2CEP: 585		T2CEP: 1006
	ACEP: 100		ACEP: 203
			RRAB: 593
			RRC: 21
			RRD: 4
Total	9575		7305

Table 1: Classification differences between data tables. See Table 8 in Appendix B for acronymns and their descriptions

Differences between object classifications of Case1 and Case2 are seen in Table 1, and as can be seen, the differences are enormous and worrying, as another column classifies 618 Cepheids as RR-Lyrae. In order to get more accurate period-luminosity relation for Cepheids, one needs to re-classify objects as have

been done in the article by V. Ripepi et al. [25]. Such re-classification goes beyond the scope of this Thesis and thus is not done here. Case1 contains column called *type2_best_sub_classification*, which gives the best subclassification[6] for T2CEP objects. Case2 does not contain such a subclassification column, but instead has a column called *best_class_score*, which describes confidence of the classification of column *best_class_name*. Values of *best_class_name* range from 0 to 1.

Another notable difference is between two columns that describe the G-band magnitude of the object, *int_average_g*[7], and *phot_g_mean_mag*[8]. As could be thought, the magnitudes for the G-band should be the same given by both columns. However, there are small differences that are caused by different computing methods, as F. Arenou, X.Luri et al. have discussed in the article "Gaia Data Release 2: Catalogue validation"[26]. The *phot_g_mean_mag* are computed directly from Gaias photometric observations and processings, while the *int_average_g* are computed by taking the average magnitude of the fitted light curve. The column *int_average_g* is designed for variable stars, as the algorithm fits the measured magnitudes into light curves, and then calculates the average magnitude of the light curve, while the *phot_g_mean_mag* is meant for non-variable stars.[26]

The average difference between the two G-band magnitudes is 0.05, with a standard deviation of 0.12, while the maximum difference is 5.83 magnitudes. Such a considerable magnitude difference leads to misleading results. Thus it is crucial to use the correct magnitude column, *int_average_g*, for variable stars.

2.2 Parallax

The equation for parallax distance —see Equation 1— works only for ideal cases, meaning that there are no measurement errors. However, as X.Luri et al. have discussed in the article *"Gaia Data Release 2, Using Gaia parallaxes"*[27], measurement errors come from different sources such as accuracy of a telescope or proper motion of the star. In some cases, the measured parallax can even be negative.

For objects that have poor fractional parallax error[9], one needs to use a more complicated form of parallax equation. Bailer-Jones has derived different methods to estimate distances from parallaxes in the article "Estimating distances from parallaxes"[28]. Methods described by Bailer-Jones uses the Bayes theorem, as Equation 1 does not behave as Gaussian because of the inverse proportional between parallax and distance.

Figure 7 shows different methods described by Bailer-Jones in the case of negative parallax and positive parallax with good fractional parallax error. The

[6] BL Herculis, W Virginis and RV Tauri are type II Cepheids.[3]

[7] "Value of the intensity-averaged magnitude in the G-band. The intensity-averaged magnitude is obtained by computing the average flux and then converting the average flux to magnitude."[23]

[8] "Mean magnitude in the G-band. This is computed from the G-band mean flux applying the magnitude zero-point in the Vega scale."[23]

[9] fractional parallax error = parallax error/parallax [28]

different methods are not discussed nor derived here because it is out of the scope of this Thesis.

The left-hand side of Figure 7 shows an example of negative parallax, and the right-hand side shows an example of positive parallax with good fractional parallax error. One sees that every method on the right-hand side agrees with each other, and Cepheids with fractional parallax error less than 0.2 are used to calibrate the Milky Way Cepheids in the coming Sections. The distance estimation from negative parallaxes is very challenging and is based on assumptions of distribution of stars in the Milky Way.[28][27] Furthermore, it is used more as a statistical method for large samples than for individual targets. The distance estimated from Figure 7 is chosen to be the mode of the posterior, as is recommended by Bailer-Jones[28] and X.Luri et al.[27].

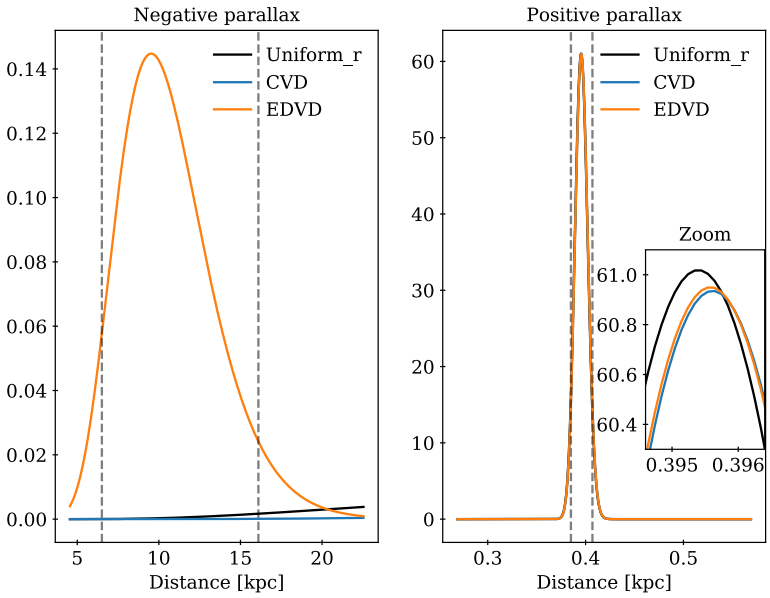

Figure 7: **Example of parallax posterior**. Left figure: Parallax -1.089mas, parallax error 0.157 and fractional parallax error -0.144. Right figure: Parallax 2.529mas, parallax error 0.042mas and fractional parallax error 0.017. Names of parallax methods follows the article by Bailer-Jones [28]. Dotted lines represent the first 5% and last 95% distance intervals.[28] See Appendix B for description of acronymns.

It has been found out that the Gaia satellite's parallax zero-point is not correct, and needs correction of +0.046mas[29][25][30]. The number of positive and negative parallaxes of the Case1 and Case2 is seen in Table 2. One can see that little over one-fifth of the parallaxes are negative.

Data set	Positive parallax	Negative parallax	NaN	Total
Case1	7327	2243	5	9575
Case2	5576	1729	0	7305

Table 2: **Table of parallaxes**. Almost one-fifth of the parallaxes are negative. For a couple of objects, the parallax could not be measured, and they are categorized as NaN.

2.3 Wesenheit magnitude

The Wesenheit magnitude is a magnitude that takes the star's color into account and is thus said to be the reddening-free magnitude of the star.[25][31] A factor that affects to reddening of the star is, among other things, the interstellar medium as at different wavelenghts, the light is affected differently by the medium.[31] The interstellar medium scatters and absorbs shorter wavelengths easier than longer wavelengths, making the background source look redder than it actually is.[3] Unfortunately, Gaia DR2 does not provide information on interstellar extinction[25], and thus extinction correction needs to be estimated from the reddening. The equation of the Wesenheit magnitude is the following:

$$W = G - \lambda(G_{BP} - G_{RP}) \tag{3}$$

Where G, G_{BP}, G_{RP} are Gaia magnitudes *int_average_g*, *int_average_bp* and *int_average_rp*[10]. Gaia DR2 also provides a column called *bp_rp* that is directly *phot_bp_mean_mag*− *phot_rp_mean_mag*[24]. Those columns are calculated using fluxes, similarly to *phot_g_mean_mag*, as described in Section 2.1. Once again, there are differences between magnitude values. However, magnitudes based on intensity-averaged method[11] gives slightly less scattered magnitude vs. period plots, and thus are used here.

The coefficient λ in Equation 3 is defined as $\lambda = A(G)/E(G_{RP} - G_{RP})$[25], where $A(G)$ is the extinction value in G-band and $E(G_{RP} - G_{RP})$ is called color excess[3]. A typical value of λ is on the order of 2 [25]. V. Ripepi et al.[25] obtained value of λ to be 1.90 by least-square fit to the type I Cepheids from Large Magellanic Cloud.

One can make a simple estimation of the value of λ by fitting a flat plane to 3D space, as seen in Figure 8. In Figure 8, one sees the 3D-plot of DCEP_F for LMC with magnitude vs. period and color index $G_{RP} - G_{RP}$. The value of λ obtained by this method is 1.63, which differs from the value obtained by Ripepi[25]. The value of λ differs obviously due to different methods used, but also due to different sample selection. The data in Figure 8 is purely what Gaia DR2[24] provides, while Ripepi et al.[25] have selected the data more precisely by doing reclassification of variables.

[10] central wavelengths: G=673nm ; G_{BP}=532nm ; G_{RP}=797nm [32]

[11] magnitude columns that begin with *int_average_*

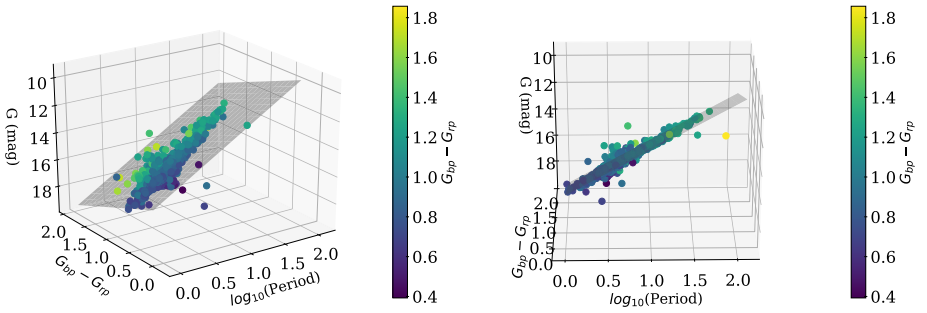

Figure 8: Period, luminosity and color index of DCEP_F from LMC plotted in 3D space, and data are taken from Case1. Color of dots represents the value of color $G_{BP} - G_{RP}$.

To study how well the values of λ are expressing the data, one can use a Chi-Squared test, which is commonly used for testing the goodness of fit.[33] Equation of the Chi-squared test is the following:

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \tag{4}$$

In order to properly compare the effect of two different λ values, PW relation of DCEP_F data points using different λ values are plotted top of each other in Figure 9. The lines seen in Figure 9 are the best fit for PW relation with λ values obtained by Ripepi et al.[25] and plane fitting. Now one can calculate Chi-squared value for both cases using Equation 4, and results are; $\chi^2_{\lambda=1.90} = 5.059$ and $\chi^2_{\lambda=1.63} = 4.857$.

The difference between χ^2 values is not huge. The value of λ obtained by fitting the plane to 3D-data space gives better results than the value obtained by Ripepi et al.[25] as could be expected. Thus the value of λ is set to be 1.63 for now on. The values of the constant α and the coefficient β —see Equation 5— for different λ values seen in Figure 9 are seen in Table 3.

	Intersection α	**Slope** β	λ	χ^2
red line	17.68 ± 0.02	-3.30 ± 0.03	1.90	5.059
blue line	17.63 ± 0.02	-3.22 ± 0.03	1.63	4.857

Table 3: Values of constant and coefficients of fitted lines from Figure 9. Lines are in the form of $mag = \alpha + \beta \log_{10}(\text{period})$, which is Equation 5.

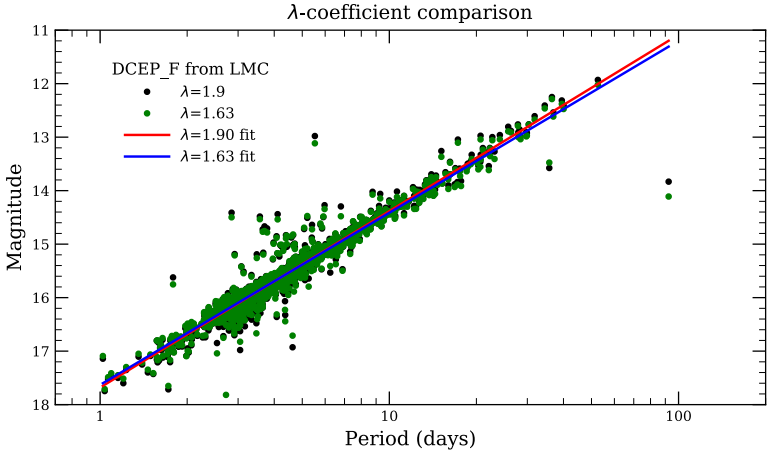

Figure 9: Comparison of different λ-coefficient values. Value 1.90 represent the result of Ripepi[25], and value 1.63 represent the result from plane fitting to 3D-data space seen in Figure 8.

3 Analysis

Before determining the Period-Luminosity (PL) relation and Period-Wesenheit relation (PW) for Milky Way (MW) Cepheids, it is done for Magellanic Clouds as their distances are well known. As described, distances to Magellanic Clouds are much larger than the size of either galaxy, and so every Cepheid within LMC or SMC can be thought to be at the same distance. That every Cepheid within the LMC or SMC can be estimated to be at the same distance, allows one to plot measured apparent magnitudes vs. periods directly, as will be seen in the next section. Later it will also be seen how the PL/PW relations obtained from MC are used to calibrate Milky Way Cepheids.

3.1 Cepheids in Magellanic Clouds from data set Case1

In Figure 10, one can see different subclassifications of Cepheid plotted as magnitude vs. period for both LMC and SMC. The subclassification of Cepheids is based on column *type_best_classification* from data set Case1. The panels (a+c) represent magnitudes from column *int_average_g*, and the panels (b+d) represent Wesenheit magnitudes, which were described in the previous section. Circles, triangles, and squares represent type I Cepheid, anomalous Cepheid, and type II Cepheid, respectively. Different colors of the symbols represent variables pulsating in the fundamental, first overtone, and multiple modes marked as _F, _10, and _M, respectively. One needs to notice that the left y-axis of Figure 10 represents apparent magnitudes, while the right y-axis represents absolute magnitudes. To calculate the absolute magnitudes using Equation 6, one needs

to know distances to LMC and SMC. The distances to the LMC and SMC are 49.59 kpc[34] and 62.1 kpc[35] respectively, where both distances have been estimated using eclipsing binaries.[34][35]

The fitted PL/PW relations can be seen in Figure 10, and those fits are made using Python3[36] library called scipy.optimize.curve_fit[37]. The PL/PW relation is in the form of

$$mag = \alpha + \beta \log_{10}(\text{period}) \tag{5}$$

Values of the constant α and coefficient β and their errors for PL/PW relations seen in Figure 10 are shown in Table 5 and 6 in Section Results 4.1.

The transition from apparent magnitudes to absolute magnitudes seen in Figure 10 is resolved using distance modulus, $m - M$, which can be written in following form[38]:

$$m - M = 5 \log_{10} \left(\frac{r}{10\text{pc}} \right) \tag{6}$$

Where m is the apparent magnitude, M is the absolute magnitude, and r is the distance to the object in parsecs. In the case of Wesenheit magnitude, one must remember to take the reddening into account, as described in Section 2.3.

3.2 Cepheids in Magellanic Clouds from data set Case2

The previous sections have shown the different classifications of Cepheids and their subclassifications, which are visible in Figure 10. However, as shown in Table 1, the classifications differ between data sets Case1 and Case2. It is essential to study and compare the cases in order to be able to select PL and PW relations justifiably.

Cepheids in the LMC and SMC based on data set Case2 can be seen in Figure 11. One needs to note that the axis scaling in Figure 11 is the same as in Figure 10 so that the differences between the data sets can be seen more clearly. Also, best fits for DCEP_F, DCEP_10, and T2CEP from Figure 10 are plotted, so that it can be confirmed that the Cepheids line up on the same fits. As described before, the data set Case2 does not provide subclassification for type II Cepheids, so in Figure 11 they are plotted with the same symbol.

The threshold limit for *best_class_score* in Figure 11 is set to be 0.7 in order to get rid of unreliable classifications. There are still misclassifications for T2CEP variables in panels (a+b), as they are mostly located in the line of DCEP_F. When comparing Figure 11 to Figure 10, one sees fewer data points, which is explained by the threshold limit of 0.7 for *best_class_score*. One also sees that there are not long periodic Cepheids in Figure 11, as there are in Figure 10. Those missing long periodic Cepheids may lead to biased data samples, as only short periodic Cepheids are represented. If one decreases the threshold limit, more data points will appear, but more falsely classified Cepheids also appear. When the threshold limit is set to be 0.8, classifications from data sets Case1 and Case2 agree for DCEP_F and DCEP_10. As mentioned, in the panels (a+b)

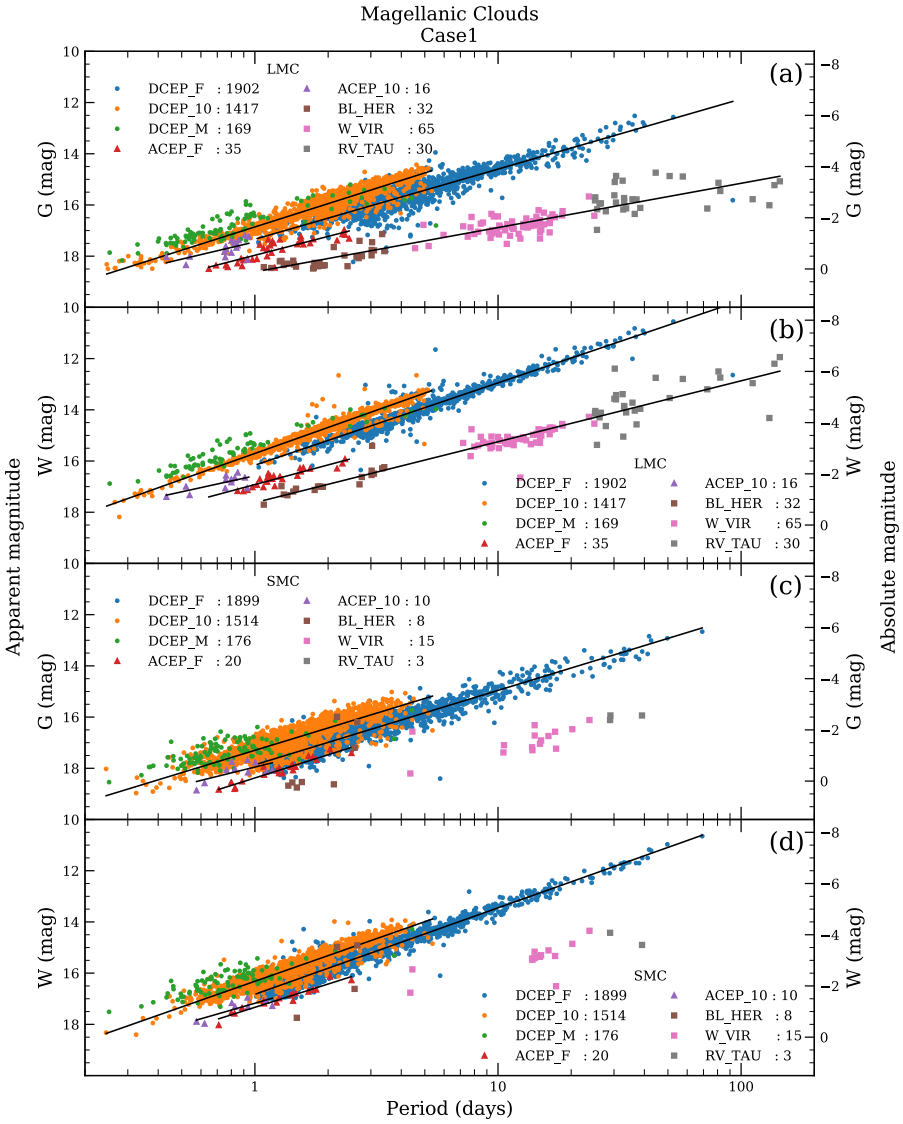

Figure 10: PL/PW relation for Large and Small Magellanic Cloud based on data set Case1. Magnitudes on left y-axis are in apparent magnitudes, and magnitudes in right y-axis are in absolute magnitudes. PL/PW relation has not been calculated for type II Cepheids in SMC because of a small number of objects and scatteration of points. Results for apparent magnitudes are in Table 5 and for absolute magnitudes in Table 6 in Section 4.1.

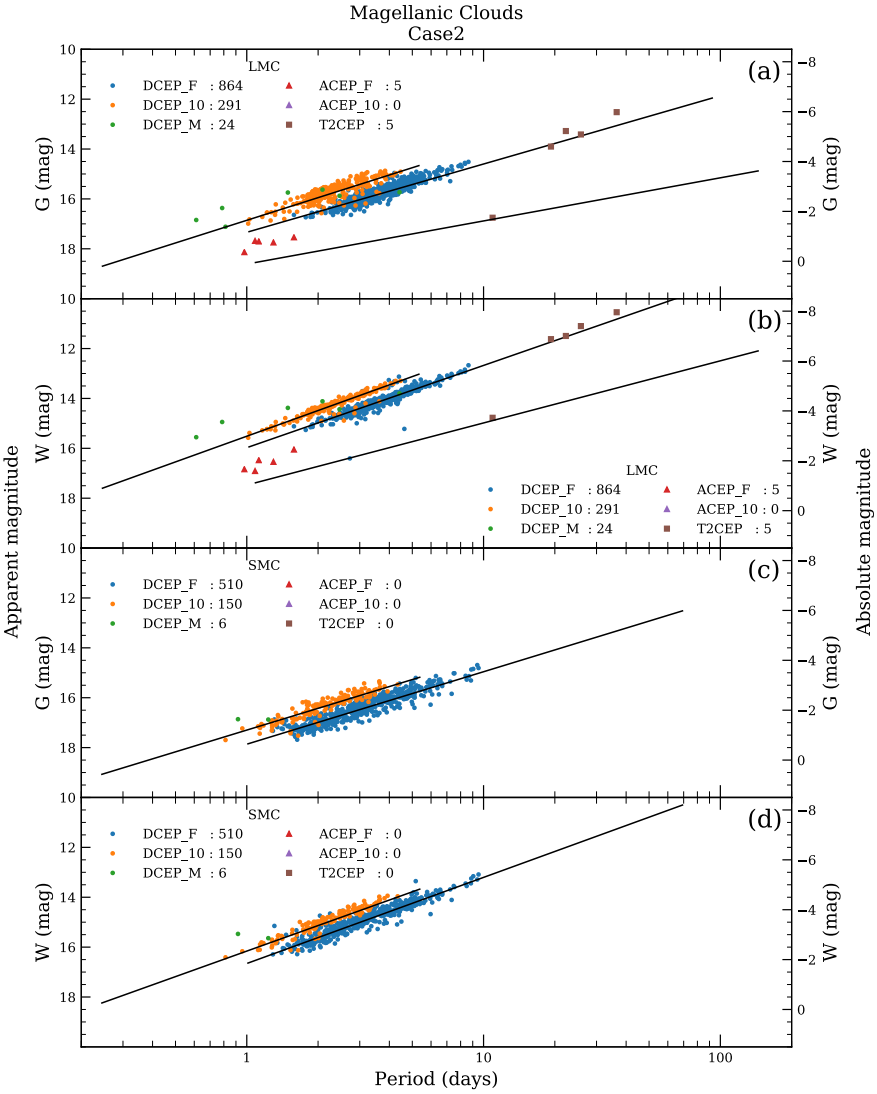

Figure 11: PL/PW relations for Large and Small Magellanic Cloud based on data set Case2. Magnitudes on left y-axis are in apparent magnitudes, andmagnitudes in right y-axis are in absolute magnitudes. Here the threshold value for *best_class_score* is set to be 0.7. The lines seen in panels are relations of DCEP_F, DCEP_10 and T2CEP from Figure 10.

Figure	Galaxy	Data set	Cepheid	Relation	Intersection α	Slope β
#11(a)	LMC	Case2	DCEP_F	PL	17.4773 ± 0.0269	-3.0684 ± 0.0480
#11(a)	LMC	Case1	DCEP_F	PL	17.3385 ± 0.0184	-2.7380 ± 0.0280
#11(a)	LMC	Case2	DCEP_10	PL	16.8004 ± 0.0422	-3.0021 ± 0.1123
#11(a)	LMC	Case1	DCEP_10	PL	16.8577 ± 0.0106	-3.0167 ± 0.0302
#11(b)	LMC	Case2	DCEP_F	PW	15.8775 ± 0.0872	-2.8506 ± 0.1533
#11(b)	LMC	Case1	DCEP_F	PW	16.1635 ± 0.0148	-3.2165 ± 0.0221
#11(b)	LMC	Case2	DCEP_10	PW	15.2246 ± 0.1045	-2.6799 ± 0.2824
#11(b)	LMC	Case1	DCEP_10	PW	15.7040 ± 0.0089	-3.3712 ± 0.0252
#11(c)	SMC	Case2	DCEP_F	PL	17.9000 ± 0.0285	-3.0465 ± 0.0548
#11(c)	SMC	Case1	DCEP_F	PL	17.8634 ± 0.0107	-2.9052 ± 0.0209
#11(c)	SMC	Case2	DCEP_10	PL	17.3876 ± 0.0414	-3.1974 ± 0.1145
#11(c)	SMC	Case1	DCEP_10	PL	17.2977 ± 0.0109	-2.8981 ± 0.0433
#11(d)	SMC	Case2	DCEP_F	PW	16.7056 ± 0.1280	-3.1302 ± 0.2421
#11(d)	SMC	Case1	DCEP_F	PW	16.8257 ± 0.0097	-3.3761 ± 0.0186
#11(d)	SMC	Case2	DCEP_10	PW	16.3285 ± 0.1718	-3.8704 ± 0.4371
#11(d)	SMC	Case1	DCEP_10	PW	16.3179 ± 0.0094	-3.3298 ± 0.0367

Table 4: Comparison of PL/PW relation values between the data sets Case1 and Case2. Relations are from Figure 11, but one need to note that relations based on the data set Case2 have not been plotted, for the sake of clarity. Errors of relations for the data set Case1 are smaller than for the data set Case2. Relations are following the form of Equation 5, $mag = \alpha + \beta \log_{10}(\text{period})$.

in Figure 11, there seem to be four T2CEP in the line of DCEP_F, and it turns out that the data set Case1 would classify those variables as DCEP_F.

Differences between PL/PW relations for DCEP_F and DCEP_10 from MC based on data sets Case1 and Case2 are seen in Table 4. When comparing values of α and β between the data set Case1 and Case2, one notices that values differ from each other quite much in some cases. Data set Case1 also provides significantly smaller errors than the data set Case2, where one source of errors for Case2 is the smaller number of data points.

Figure 12 is the panel (a) from Figure 11, with the difference that the *best_class_score* has been set to be zero. Even though setting *best_class_score* to be zero may not seem relevant, it is still the best classification guess based on the data set Case2. One needs to remember that the data set Case1 does not even have such a column, which describes the confidence of the classification. Immediately from Figure 12 can be observed that there are some severe problems within the classification of Cepheids, as over half of the type II Cepheids seems to fit with type I Cepheids. Classifications for DCEP_F and DCEP_10 seem to work well, but still, especially DCEP_F are missing long periodic variables compared to data set Case1.

Based on the observations above, it is justifiable to use PL, and PW relation from LMC and SMC obtained using the data set Case1.

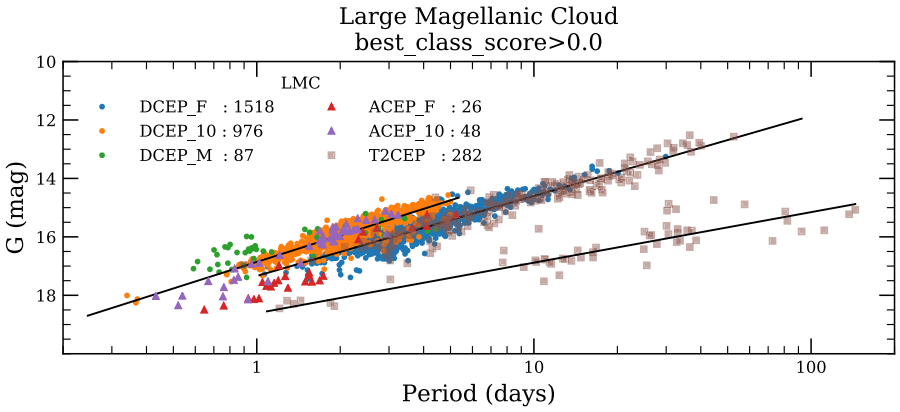

Figure 12: Relation for LMC Cepheids based on data set Case2. The threshold limit for *best_class_score* has been set to zero. Clear misclassification of Cepheids can be seen.

3.3 Milky Way Cepheids

The previous section showed that the data set Case1 gives better PL/PW relation for LMC and SMC galaxies, as seen in Table 4. Thus one could expect that the data set Case1 also gives better results for Milky Way Cepheids.

In order to calibrate PL/PW relation for MW cepheids, one needs to solve distance to the object, so that one can use Equation 6 to calculate absolute magnitude. The distance is solved using parallax methods discussed in Section 2.2. For the calibration process, one wants to be entirely sure that the distance is correct, so every parallax methods should give the same result compared to each other, as seen in the right-hand side of Figure 7. To get such a result, one needs to select such parallaxes that are positive and have the value of fractional parallax error less than 0.2[28].

Now that the distances are known, one can calculate the absolute magnitudes similarly to the previous sections. Milky Way Cepheids based on the data sets Case1 and Case2 are shown in Figure 13, and the *best_class_score* has been set to be zero for the Case2. The lines in panels (a+c) are the PL-relations, and lines in panels (b+d) are the PW-relations obtained from the LMC galaxy for DCEP_F, DCEP_10, and T2CEP variables, seen in Figure 10. The purpose of the plotted LMC relations is to make it easier to verify the classifications of MW Cepheids. The panels (a+b) are based on the classification from data set Case1, and panels (c+d) are based on the data set Case2.

One can see an unexpected 'second branch' of Cepheids in both of the Case1 and Case2. This 'second branch' might be caused by different factors such as falsely measured parallaxes, lack of proper treatment of binary objects, or misclassification of the object.[39] Figure 14 shows coordinates of the DCEP_F from Figure 13 panels (a+b), and the Cepheids from the 'second branch' are

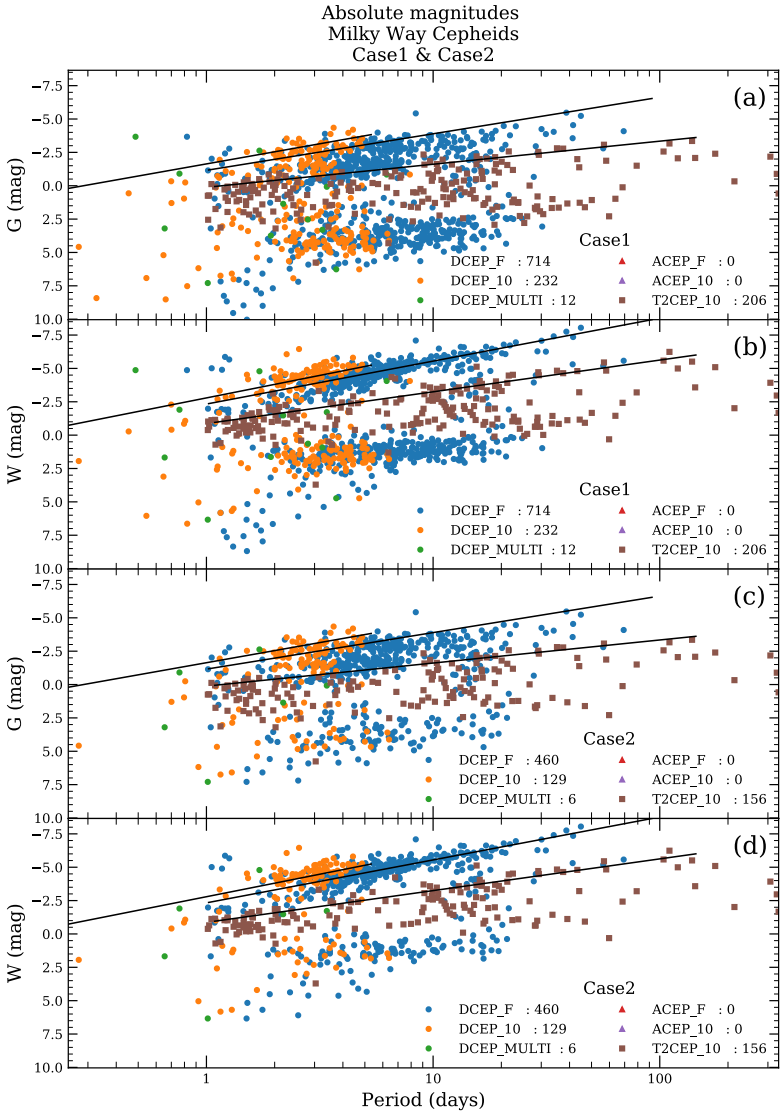

Figure 13: Absolute magnitudes of Cepheids from data sets Case1 and Case2. The *best_class_score* is set to be zero, for better comparison of the data sets. One immediately sees 'second branch' of type I Cepheids in both of the cases, which might be caused by different factors such as falsely measured parallaxes, lack of proper treatment of binary objects, or misclassification of the object.[39]

marked as red crosses. It can be seen that many of the Cepheids that are located in the 'second branch' are not in the disk plane, which supports the idea of misclassification.

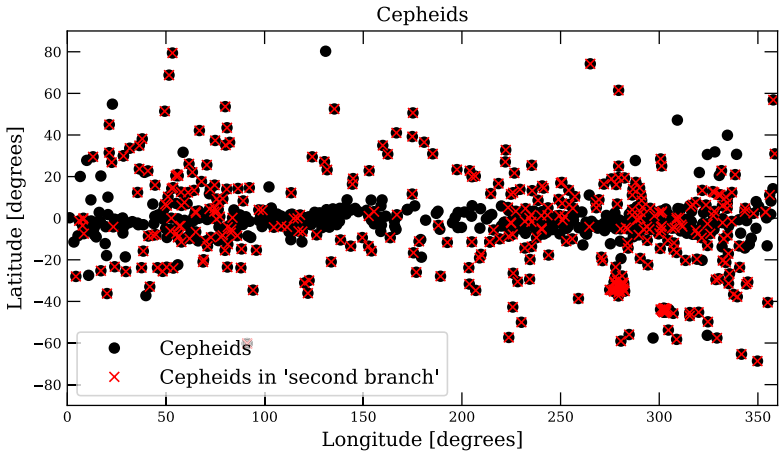

Figure 14: Cepheids in the Galactic coordinates. The red crosses represent Cepheids in the 'second branch', seen in Figure 13. It can be noticed that many of the Cepheids that are in 'second branch', are not in the disk plane. This supports the idea of misclassification.

The data set Case1 does not include a column that describes the confidence of the classification, like the data set Case2 does. Thus one can not get rid of the 'second branch' of Cepheids directly in Case1, but in Case2, one can adjust the column *best_class_score* in order to get rid of the 'second branch'. Figure 15 shows MW Cepheids based on the data set Case2 using 0.8 as the threshold value for the *best_class_score*. Using such a high threshold limit, one sees that the 'second branch' vanishes as expected. The threshold value of 0.8 is being used for now on for MW Cepheids. It is also seen that absolute G-magnitude in the panel (a), does not provide correct information on its own, which can be expected as the method for absolute G-magnitude does not take extinction into account. Cepheid variables are located in the galaxy's disk plane —see Figure 4— which means that there is much interstellar matter, dust, between observer and Cepheids causing notable extinction. As said, Gaia DR2[24] does not provide reliable information about extinction, so one needs to use the Wesenheit magnitude, and thus PW-relation.

The final thing is to verify the obtained relations. In Figure 16 (a), one sees DCEP_F variables from the LMC, SMC, and MW plotted top of each other using Wesenheit magnitudes. The small panel (b) inside the panel (a) represents the combined PW relation of the LMC and SMC galaxies, and this relation will be used to estimate distances in the next section. The panel (c) contains the same Cepheids from the panel (a) that are found from the Two Micron All Sky

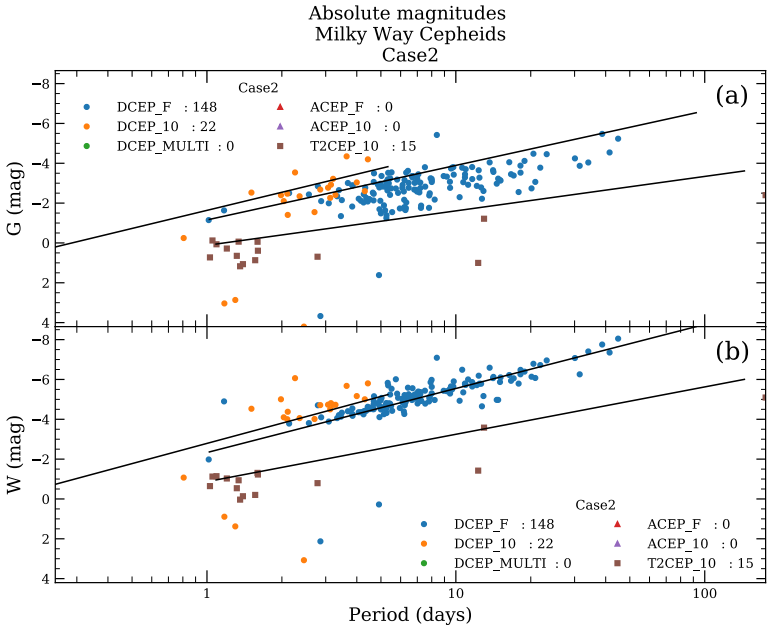

Figure 15: Absolute magnitudes of Cepheids from data set Case2. Here the threshold limit for *best_class_score* is set to be 0.8, and as seen, the 'second branch' seen in Figure 13 has disappeared.

Survey (2MASS) catalog[40]. However, instead of Wesenheit magnitude, the Cepheids are plotted in the infrared magnitudes —K_s[12]-band magnitudes—. The K_s-band magnitudes are not affected much by the extinction of dust, so they do not require extinction correction.

The PW relations in Figure 16 panel (a) agree with each other, suggesting that the calibration has been successful. The remaining thing is to confirm that the Wesenheit magnitude works as supposed. By comparing the relations in panels (a+c) and their values in Table 7, one sees that they are similar to each other. The similarity between the infrared and Wesenheit magnitudes and that the scatter around the relations is much reduced compared to pure G-band relations seen in Figure 10 panels (a+c), suggests that the Wesenheit magnitude works as supposed by removing the extinction.

Table 7 shows the results of relations from Figure 16.

[12]Central wavelength at 2.159μm[40]

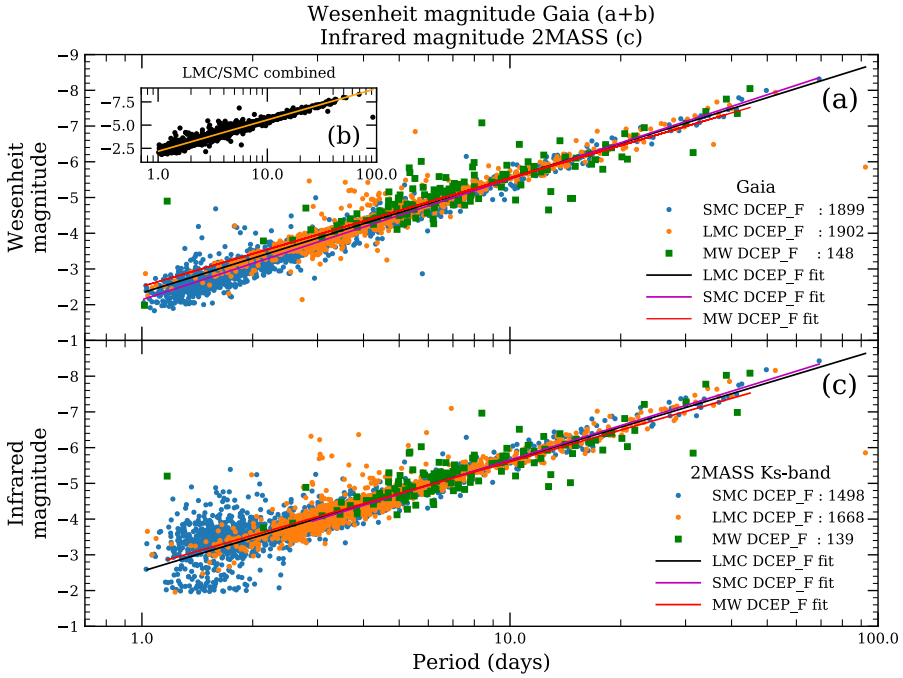

Figure 16: Panel (a) shows the PW relations for Cepheids in LMC, SMC, and MW. Panel (b) represents the combined PW relation of LMC and SMC. Panel (c) is similar to the panel (a), but instead of Wesenheit magnitudes, magnitudes are in infrared magnitudes measured by 2MASS[40]. Results for the relations are shown in Table 7.

4 Results and conclusion

4.1 Results

Tables 5 and 6 shows the values of PL/PW-relations of Figure 10. Table 5 represents the values in apparent magnitudes, and Table 6 represents the values in absolute magnitudes. When comparing results shown in Table 5 to literature values such as by Ripepi et al.[25][13] one sees that the results agree with each other. In both Tables, one also sees that the errors of α and β decreases in most cases when going from PL to PW relation.

As the results in Table 6 are in absolute magnitudes, one can compare values between the LMC and SMC relations, and they agree well with each other. Still, differences exist between the LMC and SMC relations, which might be caused by mean metallicity differences between galaxies, as theoretical models support this hypothesis.[41] Using the data provided by Gaia DR2[24], the

[13]Table 1 on page 6 of the article.

Figure	Galaxy	Cepheid	Relation	Intersection α	Slope β
#10 (a)	LMC	DCEP_F	PL	17.3385 ± 0.0184	-2.7380 ± 0.0280
#10 (a)	LMC	DCEP_10	PL	16.8577 ± 0.0106	-3.0167 ± 0.0302
#10 (a)	LMC	ACEP_F	PL	17.9649 ± 0.0463	-2.4483 ± 0.2596
#10 (a)	LMC	ACEP_10	PL	17.4429 ± 0.1127	-2.2357 ± 0.7233
#10 (a)	LMC	T2CEP	PL	18.6127 ± 0.0729	-1.7294 ± 0.0650
#10 (b)	LMC	DCEP_F	PW	16.1635 ± 0.0148	-3.2165 ± 0.0221
#10 (b)	LMC	DCEP_10	PW	15.7040 ± 0.0089	-3.3712 ± 0.0252
#10 (b)	LMC	ACEP_F	PW	16.9196 ± 0.0403	-2.5478 ± 0.2474
#10 (b)	LMC	ACEP_10	PW	16.5764 ± 0.1068	-2.0796 ± 0.6152
#10 (b)	LMC	T2CEP	PW	17.6253 ± 0.1129	-2.3784 ± 0.0935
#10 (c)	SMC	DCEP_F	PL	17.8634 ± 0.0107	-2.9052 ± 0.0209
#10 (c)	SMC	DCEP_10	PL	17.2977 ± 0.0109	-2.8981 ± 0.0433
#10 (c)	SMC	ACEP_F	PL	18.3773 ± 0.0411	-3.0105 ± 0.2350
#10 (c)	SMC	ACEP_10	PL	17.9522 ± 0.1163	-2.3488 ± 0.9863
#10 (d)	SMC	DCEP_F	PW	16.8257 ± 0.0097	-3.3761 ± 0.0186
#10 (d)	SMC	DCEP_10	PW	16.3179 ± 0.0094	-3.3298 ± 0.0367
#10 (d)	SMC	ACEP_F	PW	17.3346 ± 0.0409	-3.0183 ± 0.2211
#10 (d)	SMC	ACEP_10	PW	17.1756 ± 0.0999	-2.7349 ± 0.7982

Table 5: Values of relations from Figure 10. Relations are in apparent magnitudes, and thus representing the left y-axis of Figure 10. Relations are following the form of Equation 5, $mag = \alpha + \beta \log_{10}(\text{period})$.

Figure	Galaxy	Cepheid	Relation	Intersection α	Slope β
#10 (a)	LMC	DCEP_F	PL	-1.1384 ± 0.0184	-2.7380 ± 0.0280
#10 (a)	LMC	DCEP_10	PL	-1.6192 ± 0.0106	-3.0167 ± 0.0302
#10 (a)	LMC	ACEP_F	PL	-0.5121 ± 0.0463	-2.4483 ± 0.2596
#10 (a)	LMC	ACEP_10	PL	-1.0341 ± 0.1127	-2.2357 ± 0.7233
#10 (a)	LMC	T2CEP	PL	0.1357 ± 0.0729	-1.7294 ± 0.0650
#10 (b)	LMC	DCEP_F	PW	-2.3135 ± 0.0146	-3.2165 ± 0.0221
#10 (b)	LMC	DCEP_10	PW	-2.7730 ± 0.0089	-3.3712 ± 0.0252
#10 (b)	LMC	ACEP_F	PW	-1.5573 ± 0.0403	-2.5478 ± 0.2474
#10 (b)	LMC	ACEP_10	PW	-1.9006 ± 0.1068	-2.0796 ± 0.6152
#10 (b)	LMC	T2CEP	PW	-0.8517 ± 0.1129	-2.3784 ± 0.0935
#10 (c)	SMC	DCEP_F	PL	-1.1020 ± 0.0107	-2.9052 ± 0.0209
#10 (c)	SMC	DCEP_10	PL	-1.6678 ± 0.0109	-2.8981 ± 0.0433
#10 (c)	SMC	ACEP_F	PL	-0.5881 ± 0.0411	-3.0105 ± 0.2350
#10 (c)	SMC	ACEP_10	PL	-1.0132 ± 0.1163	-2.3488 ± 0.9863
#10 (d)	SMC	DCEP_F	PW	-2.1398 ± 0.0097	-3.3761 ± 0.0186
#10 (d)	SMC	DCEP_10	PW	-2.6475 ± 0.0094	-3.3298 ± 0.0367
#10 (d)	SMC	ACEP_F	PW	-1.6308 ± 0.0409	-3.0183 ± 0.2211
#10 (d)	SMC	ACEP_10	PW	-1.7898 ± 0.0999	-2.7349 ± 0.7982

Table 6: Values of relations from Figure 10. Relations are in absolute magnitudes, and thus representing the right y-axis of Figure 10. Relations are following the form of Equation 5, $mag = \alpha + \beta \log_{10}(\text{period})$.

mean metallicity of LMC Cepheids is -0.009, and for SMC Cepheid, it is -0.254. The difference between mean metallicities is significant, and thus differences of relations between LMC and SMC are not unexpected.

The values of PW relation from Figure 16 are shown in Table 7, and they agree well with each other. Values are only for PW relations, because as seen in Figure 15(a), the G-band magnitude on its own does not work for MW Cepheids due to extinction. When the errors in Table 7 are examined, one sees that if only MW Cepheids are used, the errors are huge compared to others. The significant errors are not surprising, because as seen in Figure 16, MW Cepheids are more scattered and lower in numbers. The smallest errors are obtained by combining the LMC and SMC relations, as seen in Table 7.

Figure	**Galaxy**	**Cepheid**	**Relation**	**Intersection α**	**Slope β**
16 (a)	LMC	DCEP_F	PW	-2.3135+-0.0146	-3.2165+-0.0221
16 (a)	SMC	DCEP_F	PW	-2.1398+-0.0097	-3.3761+-0.0186
16 (a)	MW	DCEP_F	PW	-2.5116+-0.2133	-3.0306+-0.2348
16 (b)	LMC & SMC	DCEP_F	PW	-2.1783+-0.0080	-3.3711+-0.0135
16 (c)	LMC	DCEP_F	PL (infrared)	-2.5231+-0.0177	-3.1014+-0.0270
16 (c)	SMC	DCEP_F	PL (infrared)	-2.4634+-0.0280	-3.1924+-0.0355
16 (c)	MW	DCEP_F	PL (infrared)	-2.6568+-0.2336	-2.9477+-0.2545

Table 7: Values of PW relations for DCEP_F objects from LMC, SMC, MW and LMC/SMC combined from panels (a+b) are shown. Also values for relations based on K_s-band magnitudes from the panel (c) are represented.

One can test how accurate the PW relations in Table 7 are, by calculating the distance to the known target. Example calculations are shown below. If one calculates the distance to the SMC using the PW relation obtained from the LMC, one gets the distance to be 65.6 kpc $\pm$ 10.5%. The literature distance to SMC is 62.1 kpc[35], meaning that the obtained distance is 5.6% larger than the real distance.

One can now use the results from Table 7 to calculate the absolute Wesenheit magnitude for a corresponding period. Using the notation of data columns provided by Gaia DR2[24], and the PW relation of the combined LMC/SMC, one can calculate the absolute Wesenheit magnitude for the corresponding period using Equation 5 as follows:

$$\mathrm{W}_{\text{Abs. mag.}} = (-2.1783 \pm 0.0080) + (-3.3711 \pm 0.0135)\log_{10}(\text{period})$$

Then one can calculate the distance modulus, *m-M*, while applying the definition of Wesenheit magnitude, Equation 3, as follows:

$$m - M = (int\_average\_g - 1.63(int\_average\_bp - int\_average\_rp)) - \mathrm{W}_{\text{Abs. mag.}}$$

Finally, one gets the distance using distance modulus, Equation 6, as follows:

$$r_{pw} = 10^{\frac{m-M}{5}+1}$$

One can now calculate the distances to MW Cepheids seen in Figure 16 using parallaxes and the method described above. A comparison of the distances obtained by two different methods is seen in Figure 17. The red line represents a one-to-one relation between the two distance methods.Error bars on the y-axis are 10.2% of the distances, and the error estimation was obtained by calculating the distance and its error to SMC using the combined LMC/SMC PW relation. Error bars on the x-axis represent the first 5% and last 95% distance intervals from the parallax distance posteriors, seen in Figure 7. As seen, the two methods agree well with each other as far as ∼2900pc, and after that, the scatter becomes larger.

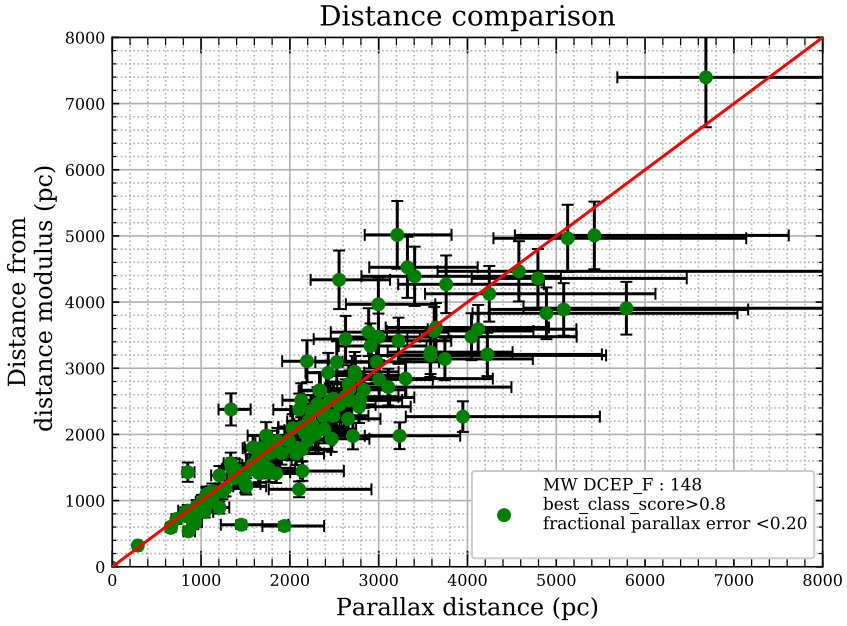

Figure 17: Distance comparison between two distance estimation methods. x-axis represents the parallax distance, and y-axis represent the distance obtained from distance modulus. Results agree well with each other up to ∼2900pc.

Possible sources for the scatter in Figure 17 are the uncertainties in parallax distances and the estimated value for the λ value in the Wesenheit magnitude. The λ value was obtained by a simple plane fitting to 3D space of magnitude, period, and color, as described in Section 2.3.

One can also calculate the distances to MW Cepheids using the pure PL relation obtained from Magellanic Clouds, but one must remember to apply the effect of extinction.

4.2 Conclusion

The main goal of this Thesis was to determine the PL/PW relations for Cepheids using the Gaia data. Table 6 shows the PL/PW relations in absolute magnitudes for different types of Cepheids in the LMC and SMC. Table 7 contains the PW relation and not the PL relation for MW Cepheids because the PL relation on its own does not work for the MW Cepheids. The reason why PL relation does not work on its own for MW Cepheids is that it does not take extinction into account, unlike the PW relation does. The comparison between the PL and PW relations for MW Cepheids are seen in Figure 15. The reason for only showing relations for DCEP_F in Table 7 is that it is the most common type of Cepheid, as seen in Table 1.

The obtained PL/PW relation values agree with literature values, and in Figure 17 are compared the distances obtained using the PW relation and parallaxes. One sees that the two distances agree with each other as far as ∼2900pc, while the scatter increases at larger distances.

One of the future goals is to test whether the PL/PW relations varies as a function of metallicity. Also, other properties of Cepheids might affect the PL/PW relation, which also requires future studies. Accurate PL/PW relations obtained using the Gaia data are used to calibrate multiple things, i.e., the value of Hubble constant H_0[25].

The secondary goal of the Thesis was to discuss inconsistencies in the Gaia data, which was done in multiple sections. Inconsistencies between data columns that should provide the same information may lead to dangerously false results, where a good example is the misclassifications between the data set Case1 and Case2. Future goal related to data selection is to be able to select correct data more accurately. The Gaia data has some more parameters that can be used for data selection, as discussed by Ripepi et al.[25]. This Thesis has shown how complex real-life data can be, even in the best missions.

"If I have seen further it is by standing on the shoulders of giants"

Sir Isaac Newton[42]

Acknowledge

Data used in this Thesis:

Gaia DR2:
"This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement."[43]

2MASS catalog
"This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation."[40]

Appendix

A: Gaia DR2 Archive quaries and 2MASS quary

Selecting data for the data set Case1

```
SELECT c.*, g.*

FROM gaiadr2.vari_cepheid as c
INNER JOIN gaiadr2.gaia_source as g

ON c.source_id = g.source_id
```

Selecting data for the data set Case2

```
SELECT c.*, g.*, r.*

FROM gaiadr2.vari_cepheid as c , gaiadr2.gaia_source as g ,
     gaiadr2.vari_classifier_result as r

WHERE c.source_id = g.source_id
AND c.source_id = r.source_id
```

Selecting data from 2MASS catalog [40]

```
raw_2mass = Gaia.launch_job_async("SELECT c.* , g.* ,r.*, \
            tmass.j_m, tmass.h_m, tmass.ks_m, tbest.angular_distance \
```

```
FROM gaiadr2.vari_cepheid as c , gaiadr2.gaia_source as g , \
     gaiadr2.vari_classifier_result as r,  \
     gaiadr2.tmass_best_neighbour AS tbest, \
     gaiadr1.tmass_original_valid AS tmass \

WHERE c.source_id = g.source_id \
    AND c.source_id = r.source_id \
    AND g.source_id = tbest.source_id  \
    AND tbest.tmass_oid = tmass.tmass_oid")

data_2mass = raw_2mass.get_results()
```

B: Acronyms

Table 8 shows used acronyms and their descriptions. Descriptions are taken from Gaia documentation[23].

Acronym	**Description**
ADQR	Astronomical Data Query Language
DR2	Data Release 2
DCEP	Classical (Delta) Cepheids
CEP	Classical (Delta) Cepheids
T2CEP	Type II Cepheids
ACEP	Anomalous Cepheids
RRAB	Fundamental-mode RR Lyrae stars
RRC	First-overtone RR Lyrae Stars
RRD	Double-mode RR Lyrae stars
BL_HER	BL Herculis
W_VIR	W Virginis
RV_TAU	RV Tauris
CVD	Constant Volume Density
EDVD	Exponentially Decreasing Volume Density
PL	Period Luminosity
PW	Period Wesenheit
MW	Milky Way
MC	Magellanic Clouds
LMC	Large Magellanic Cloud
SMC	Small Magellanic Cloud

Table 8: Table of used acronyms and their description

References

[1] James Evans, Encyclopædia Britannica, inc. Aristarchus of samos. https://www.britannica.com/biography/Aristarchus-of-Samos. [Online; accessed 19-July-2020].

[2] David McClung. Historical astronomy: Ancient greeks: Aristarchus. http://www.themcclungs.net/astronomy/people/aristarchus.html, 2014. [Online; accessed 19-July-2020].

[3] Hannu Karttunen et al. *Tähtitieteen perusteet*. Ursa, 2016.

[4] European space agency. Measuring stellar distances by parallax. https://sci.esa.int/web/gaia/-/53278-measuring-stellar-distances-by-parallax, 2013. [Online; accessed 11-June-2020].

[5] European space agency. Gaia. https://www.cosmos.esa.int/web/gaia/, 2020. [Online; accessed 19-July-2020].

[6] Swinburne University of Technology. Absolute magnitude. https://astronomy.swin.edu.au/cosmos/a/Absolute+Magnitude. [Online; accessed 16-August-2020].

[7] Chandra X-Ray Observatory. NASA/CXC/SAO. Pulsating variable stars and the hertzsprung-russell (h-r) diagram. https://chandra.harvard.edu/edu/formal/variable_stars/bg_info.html, 2015. [Online; accessed 19-July-2020].

[8] Terence Tao (UCLA). The cosmic distance ladder. https://terrytao.files.wordpress.com/2010/10/cosmic-distance-ladder.pdf, 2010. [Online; accessed 19-July-2020].

[9] Lumen Learning. Variable stars: One key to cosmic distances. https://courses.lumenlearning.com/astronomy/chapter/variable-stars-one-key-to-cosmic-distances/. [Online; accessed 19-July-2020].

[10] European Southern Observatory ESO. The first rung on the cosmic distance ladder. https://www.eso.org/public/blog/first-rung-on-cosmic-distance-ladder/, 2014. [Online; accessed 19-July-2020].

[11] NASA. Hubble space telescope. https://www.nasa.gov/mission_pages/hubble/science/star-v1.html, 2011. [Online; accessed 19-July-2020].

[12] American Association of Variable Star Observers AAVSO. The cosmic distance ladder. https://www.aavso.org/cosmic-distance-ladder, 2010. [Online; accessed 19-July-2020].

[13] Astronomy. The cosmic distance ladder: How we measure an infinite universe. https://astronomy.com/news/2019/07/the-cosmic-distance-ladder, 2019. [Online; accessed 19-July-2020].

[14] David Darling. Cosmic distance ladder. https://www.daviddarling.info/encyclopedia/C/cosmic_distance_ladder.html, 2016. [Online; accessed 19-July-2020].

[15] European Space Agency. The hipparcos space astrometry mission. https://www.cosmos.esa.int/web/hipparcos. [Online; accessed 22-July-2020].

[16] European Space Agency. The hipparcos space astrometry mission: Did you know...

[17] European space agency. Gaia overview. https://www.esa.int/Science_Exploration/Space_Science/Gaia_overview. [Online; accessed 22-July-2020].

[18] European space agency. Gaia: Mission status numbers. https://www.cosmos.esa.int/web/gaia/mission-numbers, 2020. [Online; accessed 10-June-2020].

[19] European space agency. How many stars to expect in gaia's second data release. https://sci.esa.int/web/gaia/-/60146-how-many-stars-to-expect-in-gaia-s-second-data-release, 2018. [Online; accessed 22-July-2020].

[20] European space agency. Science. https://www.cosmos.esa.int/web/gaia/science. [Online; accessed 22-July-2020].

[21] European space agency. Gaia summary. https://sci.esa.int/web/gaia/-/28820-summary. [Online; accessed 14-August-2020].

[22] European space agency. Gaia data release documentation: Processing steps. https://gea.esac.esa.int/archive/documentation/GDR2/Data$_a$$nalysis$/$chap_c u7var$/$ssec_c u7var_s os_a llsky$/$August - 2020$].

[23] European space agency. Gaia data release documentation. https://gea.esac.esa.int/archive/documentation/GDR2/, 2020. [Online; accessed 10-June-2020].

[24] European space agency. Gaia archive. https://gea.esac.esa.int/archive/, 2020. [Online; accessed 26-March-2020].

[25] V. Ripepi, R. Molinaro, I. Musella, M. Marconi, S. Leccia, and L. Eyer. Reclassification of cepheids in the gaia data release 2. *Astronomy Astrophysics*, 625:A14, Apr 2019.

[26] F. Arenou, X. Luri, C. Babusiaux, C. Fabricius, A. Helmi, T. Muraveva, A. C. Robin, F. Spoto, A. Vallenari, T. Antoja, and et al. Gaia data release 2. *Astronomy Astrophysics*, 616:A17, Aug 2018.

[27] X. Luri, A. G. A. Brown, L. M. Sarro, F. Arenou, C. A. L. Bailer-Jones, A. Castro-Ginard, J. de Bruijne, T. Prusti, C. Babusiaux, and H. E. Delgado. Gaia data release 2. *Astronomy & Astrophysics*, 616:A9, Aug 2018.

[28] Coryn A. L. Bailer-Jones. Estimating distances from parallaxes. *Publications of the Astronomical Society of the Pacific*, 127(956):994–1009, Oct 2015.

[29] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas Macri, Beatrice Bucciarelli, Mario G. Lattanzi, John W. MacKenty, J. Bradley Bowers, WeiKang Zheng, Alexei V. Filippenko, and et al. Milky way cepheid standards for measuring cosmic distances and application to gaia dr2: Implications for the hubble constant. *The Astrophysical Journal*, 861(2):126, Jul 2018.

[30] Ralph Schönrich, Paul McMillan, and Laurent Eyer. Distances and parallax bias in gaia dr2. *Monthly Notices of the Royal Astronomical Society*, 487(3):3568–3580, May 2019.

[31] Barry F. Madore, L. Freedman. Calibration of the extragalactic distance scale. https://ned.ipac.caltech.edu/level5/Cepheids/Cepheids_contents.html, 1998. [Online; accessed 3-July-2020].

[32] C. Jordi, M. Gebran, J. M. Carrasco, J. de Bruijne, H. Voss, C. Fabricius, J. Knude, A. Vallenari, R. Kohley, and A. Mora. Gaiabroad band photometry. *Astronomy Astrophysics*, 523:A48, Nov 2010.

[33] Yale University. Chi-square goodness of fit test. http://www.stat.yale.edu/Courses/1997-98/101/chigf.htm, 1997. [Online; accessed 7-July-2020].

[34] G. Pietrzyński, D. Graczyk, A. Gallenne, W. Gieren, I. B. Thompson, B. Pilecki, P. Karczmarek, M. Górski, K. Suchomska, M. Taormina, and et al. A distance to the large magellanic cloud that is precise to one per cent. *Nature*, 567(7747):200–203, Mar 2019.

[35] Dariusz Graczyk, Grzegorz Pietrzyński, Ian B. Thompson, Wolfgang Gieren, Bogumił Pilecki, Piotr Konorski, Andrzej Udalski, Igor Soszyński, Sandro Villanova, Marek Górski, and et al. The araucaria project. the distance to the small magellanic cloud from late-type eclipsing binaries. *The Astrophysical Journal*, 780(1):59, Dec 2013.

[36] Python Software Foundation . Python3. https://www.python.org/doc/, 2020. [Online; accessed 3-July-2020].

[37] The SciPy community. Scipy optimize curve_fit. https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html, 2020. [Online; accessed 3-July-2020].

[38] Swinburne University of Technology. Distance modulus. https://astronomy.swin.edu.au/cosmos/d/Distance+Modulus. [Online; accessed 7-July-2020].

[39] G. Clementini, V. Ripepi, R. Molinaro, A. Garofalo, T. Muraveva, L. Rimoldini, L. P. Guy, G. Jevardat de Fombelle, K. Nienartowicz, O. Marchal, and et al. Gaia data release 2. *Astronomy Astrophysics*, 622:A60, Jan 2019.

[40] NASA/IPAC. Two micron all sky survay 2mass. https://irsa.ipac.caltech.edu/Missions/2mass.html, 2020. [Online; accessed 8-July-2020].

[41] Allan Sandage and Gustav A. Tammann. Absolute magnitude calibrations of population i and ii cepheids and other pulsating variables in the instability strip of the hertzsprung-russell diagram. *Annual Review of Astronomy and Astrophysics*, 44(1):93–140, 2006.

[42] Historical Society of Pennsylvania. Newton, i., robert hooke. isaac newton letter to robert hooke, 1675. https://discover.hsp.org/Record/dc-9792/Descriptiontabnav, 2020. [Online; accessed 12-July-2020].

[43] European space agency. Gaia citation. https://gea.esac.esa.int/archive/documentation/GDR2/Miscellaneous /sec_credit_and_citation_instructions/, 2020. [Online; accessed 12-July-2020].